Bradley Gerber DATA-412
Lily Stockbridge DATA-612
Owen Albanese DATA-612

Rural vs Urban Discrepancies Across Regions

We propose to analyze the growth of urban populations across the world between the years of 2005 and 2018. It is well known that since the turn of this millennium, people have coalesced around urban population centers around the world in far greater numbers than ever before. We want to visualize that change and attempt to find connections within the data which could illuminate some of the larger trends within this migration.

The datasets we've come up with would allow comparisons between regions, over time, and against their GDP.

- UN data on urban vs rural population by country between 2005 and 2018
- World Bank data on the percentages of national populations living in a city with over 1 million people between 1990 and 2021
- UN data on percentage of national populations categorized as urban between 1950 and 2050
- OECD data on GDP per capita between urban and rural regions within nations between 2001 and 2020

Collected by the UN, World Bank, and OECD, each of these would offer significant potential to find macrotrends in the migration from rural to urban populations. We will attempt to clean whichever dataset we choose so that null values are removed and inconsistencies are properly identified. Then, it will be visualized over a range of factors to determine which variables contribute to a unique finding. We plan to utilize line plots grouped by region to find systematic differences between global development as well as histograms to gain an understanding of urbanization on a global scale. As for the statistical analysis, the DATA-612 team-members will conduct a T-test (and/or possibly more depending on results) to see if the changes in populations are significant over time.

Responsibilities for this project will be split evenly between all three members such that all members substantially contribute to all aspects of the final project (besides the DATA-612 differentiation).

Project milestones will be discussed at least weekly - whether online or in-person - with additional meetings scheduled whenever necessary. Primary collaboration will occur within Google Docs and text channels.

We have agreed upon the following timeline for key steps of the project:

Milestone	Start Date	Due Date
Gathering Data	10/20	10/24

Cleaning Data	10/24	10/31
Conducting the analysis: EDA and Statistical	10/31	11/13 (provided in progress report)
Producing Report	11/14	11/27
Producing the Group Presentation Document	11/28	12/4
Rehearsing the Presentation	12/5	12/11